

Statistical mechanics approach to a reinforcement learning model with memory

Adam Lipowski,¹ Krzysztof Gontarek,¹ and Marcel Ausloos²

¹*Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland*

²*GRAPES, University of Liège, B-4000 Liège, Belgium*

We introduce a two-player model of reinforcement learning with memory. Past actions of an iterated game are stored in a memory and used to determine player's next action. To examine the behaviour of the model some approximate methods are used and confronted against numerical simulations and exact master equation. When the length of memory of players increases to infinity the model undergoes an absorbing-state phase transition. Performance of examined strategies is checked in the prisoner's dilemma game. It turns out that it is advantageous to have a large memory in symmetric games, but it is better to have a short memory in asymmetric ones.

I. INTRODUCTION

Game theory plays an increasingly important role in many disciplines such as sociology, economy, computer sciences or even philosophy [1]. Providing a firm mathematical basis, this theory stimulates development of quantitative methods to study general aspects of conflicts, social dilemmas, or cooperation. At the simplest level such situations can be described in terms of a two-person game with two choices. In the celebrated example of such a game, the Prisoner's Dilemma, these choices are called cooperate (C) and defect (D). The single Nash equilibrium, where both players defect, is not Pareto optimal and in the iterated version of this game players might have some incentives to cooperate. However, finding an efficient strategy even for such a simple game is highly nontrivial albeit exciting task, as evidenced by the popularity of Axelrod's tournaments [2]. These tournaments had the unquestionable winner - the strategy tit-for-tat. Playing in a given round what an opponent played in the previous round, the strategy tit-for-tat is a surprising match of effectiveness as well as simplicity. Later on various strategies were examined: deterministic, stochastic, or evolving in a way that mimic biological evolution. It was also shown that some strategies perform better than the strategy tit-for-tat, as an example one can mention the strategy called *win-stay loose-shift* [3]. In an interesting class of some other strategies previous actions are stored in the memory and used to determine future actions. However, since the number of possible previous actions increases exponentially fast with the length of memory and a strategy has to encode the response for each of such possibilities, the length of memory has to be very short [4]. Such a short memory cannot detect possible longer-term patterns or trends in the actions of the opponent.

Actually, the problem of devising an efficient strategy that would use the past experience to choose or avoid some actions is of much wider applicability, and is known as reinforcement learning. Intensive research in this field resulted in a number of models [5], but mathematical foundations and analytical insight into their behaviour seems to be less developed. Much of the theory of the reinforcement learning is based on the Markov Decision

Processes where it is assumed that the player environment is stationary [6]. Extension of this essentially single-player problem to the case of two or more players is more difficult but some attempts have been already made [7]. Urn models [8] and various buyers-sellers models [9] were also examined in the context of reinforcement learning.

In most of the reinforcement learning models [10, 11] past experience is memorized only as an accumulated payoff. Although this is an important ingredient, storing the entire sequence of past actions can potentially be more useful in devising efficient strategies. To get a preliminary insight into such an approach, in the present paper we introduce a model of an iterated game between two players. A player stores in its memory the past actions of an opponent and uses this information to determine probability of its next action. We formulate approximate methods to describe the behaviour of our model and confront them against numerical simulations and exact master equation. Let us notice that numerical simulations are the main and often the only tool in the study of reinforcement learning models. The possibility to use analytical and sometimes even exact approaches such as those used in the present paper seems to be a rare exception. Our calculations show that when the length of memory increases to infinity, a transition between different regimes of our model takes place, that is analogous to an absorbing-state phase transition [12]. Similar phase transitions might exist in spatially extended, multi-agent systems [13], however in the introduced two-player model this transition has a much different nature, namely it takes place only in the space of memory configurations.

II. A REINFORCEMENT LEARNING MODEL WITH MEMORY

In our model we consider a pair of players playing repeatedly a game like e.g., the prisoner's dilemma. A player i ($i = 1, 2$) is equipped with a memory of length l_i , where it sequentially stores the last l_i decisions made by its opponent. For simplicity let us consider a game with two decisions that we denote as C and D. An example that illustrates a memory change in a single round of a game is shown in Fig. 1 (we will mostly examine the

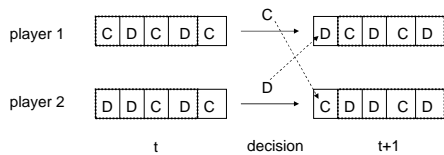


FIG. 1: Memory change during a single round of a game with two players with memories of length $l = 5$. The first player shifts all memory cells to the right (removing the rightmost element) and puts the last decision (D) of the second player at the left end. Analogous change takes place in the memory of the second player

symmetrical case where $l_1 = l_2$, and the index i denoting the player will be thus dropped).

A player uses the information in its memory to evaluate the opponent's behaviour and to calculate probabilities of making its own decisions. Having in mind a possible application to the prisoner's dilemma we make the eagerness to cooperate of a player to be dependent on the frequency of cooperation of its opponent. More specifically, we assume that the probability p_t for a player to play C at the time t is given by

$$p_t = 1 - ae^{-bn_t/l}, \quad (1)$$

where n_t is the number of C's in player's memory at time t while $a > 0, b > 0$ are some additional parameters. In principle a can take any value such that $0 < a \leq 1$ but numerical calculations presented below were made only for $a = 1$ that left us with only two control parameters, namely b and l , that determine the behaviour of the model. For $a = 1$ the model has an interesting absorbing state: provided that both players have $n_t = 0$ they both have $p_t = 0$ and thus they will be forever trapped in this (noncooperative) state. As we will see, this feature in the limit $l \rightarrow \infty$ leads to a kind of phase transition (already in the case of two players).

The content of the memory in principle might provide much more valuable information on the opponent behaviour than Eq. (1) which is only one of the simplest possibilities. As we already mentioned, our choice of the cooperation probability(1) was motivated by the Prisoner's Dilemma but of course for other games different expressions might be more suitable. Moreover, more sophisticated expressions, for example based on some trends in the distribution of C's, might lead to more efficient strategies but such a possibility is not explored in the present paper.

Let us also notice, that in our approach the memory of a player stores the sequence of past actions of length l (and that information is used to calculate the probability of cooperation). We do not store the response to each possible past sequence of actions (as e.g., in [4]) and that is why memory requirements in our model increase only linearly with l and not exponentially.

A. Mean-value approximation

Despite a simple formulation the analysis of the model is not entirely straightforward. This is mainly because the probability p_t is actually a random variable that depends on the dynamically determined content of a player's memory. However, some simple arguments can be used to determine the evolution of p_t at least for large l . Indeed, in such a case one might expect that fluctuations of n_t/l are negligible and it might be replaced in Eq. (1) with its mean value. Since at time t the coefficient n_t of player (1) equals to the number of C's made by its opponent (2) during l previous steps we obtain the following expression for its mean value

$$\langle n_t^{(1)} \rangle = \sum_{k=1}^l p_{t-k}^{(2)}, \quad (2)$$

where the upper indices denote the players. Under such an assumption we obtain that the evolution of probabilities $p_t^{(1,2)}$ is given by the following equations

$$p_t^{(1,2)} = 1 - \exp\left(\frac{-b}{l} \sum_{k=1}^l p_{t-k}^{(2,1)}\right) \quad t = l+1, l+2, \dots \quad (3)$$

In Eq. (3) we assume that both players are characterized by the same values of b and l , but generalization to the case where these parameters are different is straightforward. To iterate Eq. (3) we have to specify $2l$ initial values. For the symmetric choice

$$p_t^{(1)} = p_t^{(2)}, \quad t = 1, 2, \dots, l, \quad (4)$$

we obtain symmetric solutions (i.e., with Eq (4) being satisfied for any t). In such a case the upper indices in Eq. (3) can be dropped.

For large l the mean-value approximation (3) is quite accurate. Indeed, numerical calculations show that already for $l = 40$ this approximation is in very good agreement with Monte Carlo simulations (Fig. 2). However, for smaller l a clear discrepancy can be seen.

Provided that in the limit $t \rightarrow \infty$ the system reaches a steady state ($p_t = p$), in the symmetric case we obtain

$$p = 1 - \exp(-bp). \quad (5)$$

Elementary analysis show that for $b \leq 1$ the only solution of (5) is $p = 0$ and for $b > 1$ there is also an additional positive solution. Such a behaviour typically describes a phase transition at the mean-field level, but further discussion of this point will be presented at the end of this section.

B. Independent-decisions approximation

As we already mentioned, the mean-value approximation (3) neglects fluctuations of n_t around its mean value.

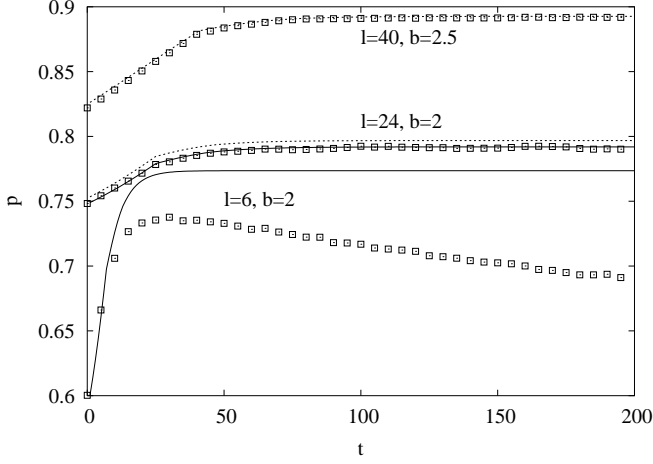


FIG. 2: The cooperation probability p as a function of time t . The dashed lines correspond to the mean-value approximation (3) while the continuous line shows the solution of independent-decisions approximations (7). Simulation data (\square) are averages over 10^4 independent runs. For $l = 24$ simulations and independent-decisions approximation (7) are in a very good agreement while mean-value approximation (3) slightly differs. For $l = 40$ calculations using (7) are not feasible but for such a large l a satisfactory description is obtained using the mean-value approximation (3). Calculations for $l = 6$ shows that independent-decisions approximation deviates from simulations. Results of approx. (3) are not presented but in this case they differ even more from simulation data. The decrease of p as seen in the simulation data is due to the small probability of entering an absorbing state (no cooperation). On the other hand, approximations (3) as well as (7) predict that for $t \rightarrow \infty$ the probability p tends to a positive value. For $l = 24$ and 40 as initial conditions we took (symmetric case) $p_t = 0.7$, $t = 1, 2, \dots, l$ and for $l = 6$ we used $p_t = 0.5$. Initial conditions in Monte Carlo simulations corresponded to these values.

In this subsection we try to take them into account. Let us notice that a player with memory length l can be in one of the 2^l configurations (*conf*). Provided that we can calculate probability p_{conf} of being in such a configuration (at time t), we can write

$$p_t = \sum_{conf} \left[1 - \exp \left(-\frac{bn(conf)}{l} \right) \right] p_{conf}, \quad (6)$$

where $n(conf)$ is the number of C 's in a given configuration *conf* and the summation is over all 2^l configurations; indices of players are temporarily omitted. But for a given configuration we know its sequence of C 's and D 's and thus its history. For example, if at time t a memory of a player (with $l = 3$) contains CDD it means that at time $t - 1$ its opponent played C and at time $t - 2$ and $t - 3$ played D (we use the convention that most recent elements are on the left side). Assuming that such actions are independent, in the above example the probability of the occurrence of this sequence might be written as $p_t(1 - p_{t-1})(1 - p_{t-2})$. Writing p_{conf} in such a product

form for arbitrary l , Eq. (6) can be written as

$$p_t = \sum_{\{E_k\}} \left[1 - \exp \left(-\frac{bn(\{E_k\})}{l} \right) \right] \prod_{k=1}^l f_{t-k}(E_k), \quad (7)$$

where the summation in Eq. (7) is over all 2^l configurations (sequences) $\{E_k\}$ where $E_k = C$ or D and $k = 1, \dots, l$. Moreover, $n(\{E_k\})$ equals the number of C 's in a given sequence and

$$f_{t-k}(E_k) = \begin{cases} p_{t-k} & \text{for } E_k = C \\ 1 - p_{t-k} & \text{for } E_k = D \end{cases} \quad (8)$$

For $l = 2$, Eq. (7) can be written as

$$\begin{aligned} p_t^{(1,2)} &= p_{t-1}^{(2,1)} p_{t-2}^{(2,1)} r_2 + p_{t-1}^{(2,1)} (1 - p_{t-2}^{(2,1)}) r_1 + \\ &+ (1 - p_{t-1}^{(2,1)}) p_{t-2}^{(2,1)} r_1 + \\ &+ (1 - p_{t-1}^{(2,1)}) (1 - p_{t-2}^{(2,1)}) r_0, \end{aligned} \quad (9)$$

where $r_k = 1 - \exp(-bk/2)$.

The number of terms in the sum of Eq. (7) increases exponentially with l , but numerically one can handle calculations up to $l = 24 \sim 28$. Solution of Eq. (7) is in much better agreement with simulations than the mean-value approximation (3). For example for $l = 24$ and $b = 2$ it essentially overlaps with simulations, while (3) clearly differs (Fig.2).

Despite an excellent agreement seen in this case, the scheme (7) is not exact. As we already mentioned, this is because the product form of the probability p_{conf} is based on the assumption that decisions at time $t - 1$, $t - 2, \dots, t - l$ are independent, while in fact they are not. For smaller values of l the (increasing in time) difference with simulation data might be quite large (Fig.2).

C. Master equation

In this subsection we present the exact master equation of this system. This equation directly follows from the stochastic rules of the model and describes the evolution of probabilities of the system being in a given state. Let us notice that a state of the system is given by specifying the memory content of both agents. In the following we present the explicit form of this equation only in the case $l = 2$, but an extension to larger l is straightforward but tedious. We denote the occupation probability of being at time t in the state where the first player has in its memory the values E, F and the second one has G and H as $p_t^{EF, GH}$. Assuming that the parameters b and l are the same for both players and that symmetric initial conditions are used

$$p_t^{EF, GH} = p_t^{GH, EF}, \quad t = 0 \quad (10)$$

enables us to reduce the number of equations from 16 to 10. The resulting equations preserve the symmetry (10)

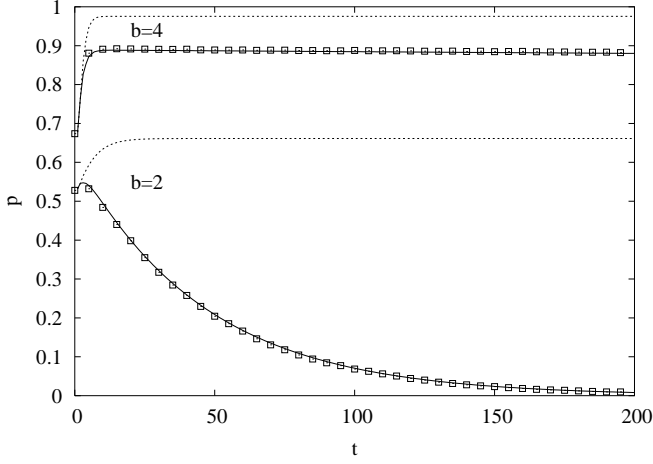


FIG. 3: The cooperation probability as a function of time t for two players with $l = 2$. Exact master equation solution (11)-(12) (solid line) is in perfect agreement with simulations (\square) and deviates from the independent-decisions approximation (9) (dotted line).

for any t and are the same for each of the players. The master equation of our model for $t = 1, 2, \dots$ takes the following form

$$\begin{aligned}
 p_t^{CC,CC} &= p_{t-1}^{CC,CC} r_2^2 + 2p_{t-1}^{CC,CD} r_2 r_1 + \\
 &\quad + p_{t-1}^{CD,CD} r_1^2 \\
 p_t^{CC,CD} &= p_{t-1}^{CC,CC} r_2 r_1 + p_{t-1}^{CD,DC} r_1^2 \\
 p_t^{CC,DC} &= p_{t-1}^{CC,CC} r_2 (1 - r_2) + p_{t-1}^{CD,CD} r_1 (1 - r_1) + \\
 &\quad + p_{t-1}^{CC,CD} (r_1 + r_2 - 2r_1 r_2) \\
 p_t^{CC,DD} &= p_{t-1}^{CC,DC} r_1 (1 - r_2) + p_{t-1}^{CD,DC} r_1 (1 - r_1) \\
 p_t^{CD,DC} &= p_{t-1}^{CC,DC} r_2 (1 - r_1) + p_{t-1}^{CC,DD} r_2 + \\
 &\quad + p_{t-1}^{CD,DC} r_1 (1 - r_1) + p_{t-1}^{CD,DD} r_1 \\
 p_t^{DC,DC} &= p_{t-1}^{CC,CC} (1 - r_2)^2 + p_{t-1}^{CC,CD} (1 - r_2) (1 - r_1) + \\
 &\quad + p_{t-1}^{CC,CD} (1 - r_2) (1 - r_1) + p_{t-1}^{CD,CD} (1 - r_1)^2 \\
 p_t^{CD,CD} &= p_{t-1}^{DC,DC} r_1^2 \\
 p_t^{DC,DD} &= p_{t-1}^{CD,DD} (1 - r_1) + p_{t-1}^{CC,DD} (1 - r_2) + \\
 &\quad + p_{t-1}^{CD,DC} (1 - r_1)^2 + p_{t-1}^{CC,DC} (1 - r_1) (1 - r_2) \\
 p_t^{CD,DD} &= p_{t-1}^{DC,DD} r_1 + p_{t-1}^{DC,DC} r_1 (1 - r_1) \\
 p_t^{DD,DD} &= p_{t-1}^{DD,DD} + 2p_{t-1}^{DC,DD} (1 - r_1) + \\
 &\quad + p_{t-1}^{DC,DC} (1 - r_1)^2.
 \end{aligned} \tag{11}$$

Iterating Eq. (11) one can calculate all occupation probabilities $p_t^{EF,GH}$. The result can be used to obtain the probability of cooperating at time t

$$\begin{aligned}
 p_t &= p_{t+1}^{CC,CC} + p_{t+1}^{CC,DC} + 2p_{t+1}^{CC,CD} + p_{t+1}^{CC,DD} + \\
 &\quad + p_{t+1}^{CD,CD} + p_{t+1}^{CD,DC} + p_{t+1}^{CD,DD}.
 \end{aligned} \tag{12}$$

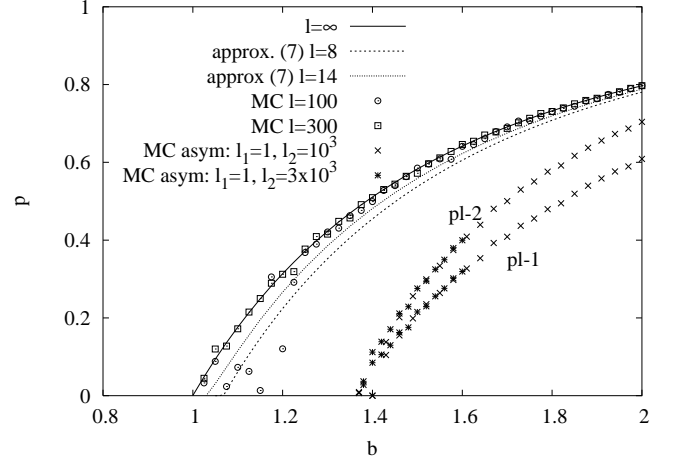


FIG. 4: The steady-state cooperation probability p as a function of b . The independent-decision approximation (7) for increasing l converges to the mean-value approximation (5) that in the limit $l = \infty$ presumably becomes exact. In the asymmetric case the cooperation probability of each player is different. The first player (pl-1) has the memory length $l_1 = 1$ and the second player (pl-2) has $l_2 = 10^3$ or $3 \cdot 10^3$.

For $b = 2$ and 4 the numerical results are presented in Fig. 3. One can see that they are in perfect agreement with simulations. Let us notice that for $b = 2$ after a small initial increase, the cooperation probability p_t decreases in time. This is an expected feature and is caused by the existence of the absorbing state DD,DD. Of course, the equations (11) reflect this fact: the probability $p_{t-1}^{DD,DD}$ enters only the last equation, namely that describing the evolution of $p_t^{DD,DD}$ (in other words, none of the states can be reached from this state). Although on a larger time scale p_t would decrease also for $b = 4$, on the examined time scale it seems to saturate at a positive value. Solutions (i.e., p_t) obtained from the independent-decisions approximation as well as mean-value approximation saturates at some positive values in the limit $t \rightarrow \infty$ and thus approximately correspond to such quasi-stationary states.

The (quasi-)stationary behaviour of the model is presented in Fig. 4. Provided that b is large enough the players remain in the cooperative phase; otherwise they enter the absorbing (noncooperative) state. However, for finite memory length l the cooperative state is only a transient state, and after a sufficiently large time an absorbing state will be reached. Thus, strictly speaking, a phase transition between cooperative and noncooperative regimes takes place only in the limit $l \rightarrow \infty$. In this limit the mean-value approximation (5) correctly describes the behaviour of the model. Simulations agree with (5), but to obtain good agreement for b close to the transition point value $b = 1$, the length of memory l should be large.

We have also examined the nonsymmetric (with respect to the memory length) case, where the first player

player-1 \ player-2	C	D
C	(3,3)	(0,5)
D	(5,0)	(1,1)

FIG. 5: The payoff matrix of the prisoner's dilemma game used in the calculations presented in Figs. 6-7. The first and the second number of a pair in a given cell denotes payoff of the first and second player, respectively.

has the memory of finite length l_1 and the length of the memory of the second player l_2 diverges. Simulations for $l_1 = 1$ and $l_2 = 10^3$ and $3 \cdot 10^3$ show that in this case there is also a phase transition (Fig. 4) but at a larger value of b than in the symmetric case (apparently, fluctuations due to the short memory of the first player ease the approach of an absorbing state). Results for larger values of l_1 (not presented) show that this transition approaches the phase transition in the symmetric case.

The phase transition that is shown in Fig. 4 is an example of an absorbing-state phase transition with cooperative and noncooperative phases corresponding to active and absorbing phases, respectively [12]. Such transitions appear also for some models of Prisoner's Dilemma (or other games) in spatially extended systems [13], i.e., the phase transition appears in the limit when the number of players increases to infinity. In the present model the nature of this transition is much different: the number of players remains finite (and equal to two) but the length of memory diverges.

III. PRISONER'S DILEMMA

In this section we examine our players in an explicit example of the Prisoner's Dilemma with the typically used payoff matrix that is shown in Fig. 5. Results of the calculations of the time dependence of the average payoff are presented in Figs. 6-7. Simulations in the symmetric case (Fig. 6) show that the larger the memory length l , the larger the payoff. In the asymmetric case (Fig. 7) the shorter-memory player for large t has larger payoff, but initially it might have the smaller payoff than the longer-memory player. In simulations shown in Figs. 6-7 the memory length was rather short and the model relatively quickly enters the absorbing (noncooperative) state. That is why the average payoff converges asymptotically to unity. Although this is not shown, such a behaviour was seen also in the asymmetric case, but on a larger time scale than that presented in Fig. 7.

Using solely the results shown in Figs. 6-7 it is difficult to predict what are the parameters (l , b) of the best (i.e., accumulating the largest payoff) player. This is because

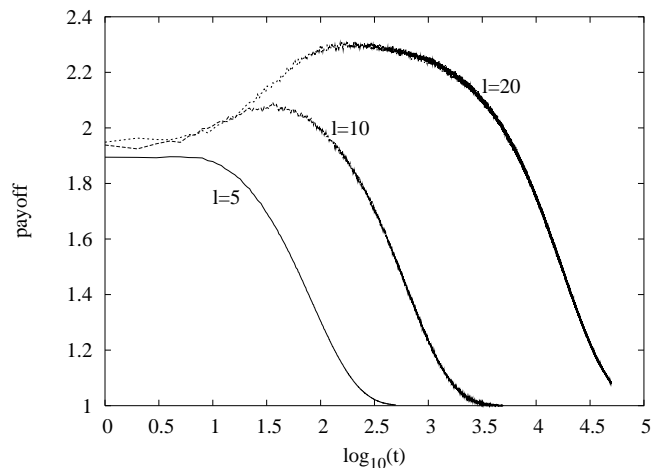


FIG. 6: The time evolution of the average payoff in the symmetric case ($l_1 = l_2 = l$) for $b = 1.5$ and several values of l . Results are averages over 10^5 independent runs. As an initial state each player at each cell of its memories has C or D with probabilities 0.3 and 0.7, respectively.

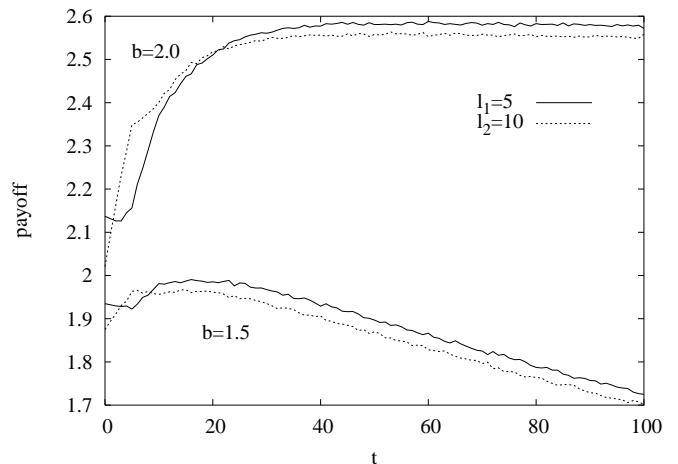


FIG. 7: The time evolution of the average payoff in the asymmetric case $l_1 = 5$ and $l_2 = 10$. Results are averages over 10^5 independent runs. As an initial state each player at each cell of its memories has C or D with probabilities 0.3 and 0.7, respectively.

the performance of a given player depends on the parameters of the opponent, number of rounds or even the initial content of the memory. And already the length of memory alone results in conflicting properties: it pays off to have a large memory in symmetric games (Fig. 6), but it is better to have a short memory in asymmetric ones (Fig. 7). It would be thus interesting to perform Axelrod's type tournament that would make the evolutionary selection of the winner, where the accumulated payoff of each player would determine its fitness. Particularly interesting might be to examine a spatially extended version of such a tournament, where opponents of a given player would be only its neighbouring sites. In such a

tournament one can check for example whether spatial effects modify the nature (i.e., universality class) of the absorbing-state phase transition. And of course, it would be interesting to check whether in such an ensemble of players the strategy tit-for-tat, that in our model is obtained for $l = 1$ and $b \rightarrow \infty$, will be again invincible.

As a further extension one can consider playing multi-decision games. In such a case an additional group structure might appear and examination of the nature of cooperation becomes much more subtle [14].

IV. CONCLUSIONS

In the present paper we have introduced a reinforcement learning model with memory and have analysed it using approximate methods, numerical simulations and exact master equation. In the limit when the length of memory becomes infinite the model has an absorbing-state phase transition. The objective of the paper was to develop general approaches (such as approximate descriptions or master-equation analysis) to study such models, and that is why rather a simple and motivated by the Prisoner Dilemma form (1) of the cooperation probability, was used. In some particular games more complicated functions might prove more efficient. One can also consider storing in player's memory some additional information concerning, e.g., players own moves. Perhaps analytical approaches, that we used in some simple exam-

ples, can be adapted to such more complicated problems as well.

We also suggested that it would be desirable to perform Axelrod's type tournament for players with memory (as in our work), but in addition equipped with some evolutionary abilities [15]. Such a tournament would allow us to examine the coexistence of learning and evolution that is an interesting subject on its own. Better learning abilities might influence the survival and thus direct the evolution via the so-called Baldwin effect [16]. Some connections between learning and evolution were already examined also in the game-theory setup [17, 18]. For the present model a detailed insight at least into learning processes is available and coupling them with evolutionary processes might lead to some interesting results in this field.

Finally, let us notice that decision making based on the content of memory seems to be connected with the psychophysical relation between response and stimulus. Early attempts to express such a relation in mathematical terms lead to the so-called Weber-Fechner law [19]. Despite some works that reproduce this type of law [20], further research, perhaps using models similar to those described in the present paper, would be desirable.

Acknowledgments: We gratefully acknowledge access to the computing facilities at Poznań Supercomputing and Networking Center. A.L. and M.A were supported by the bilateral agreement between University of Liège and Adam Mickiewicz University.

-
- [1] D. Fudenberg and J. Tirole, *Game Theory*, (MIT Press, Cambridge, Massachusetts, 1991).
 - [2] R. Axelrod, *The Evolution of Cooperation* (Basic Books, New York, 1984).
 - [3] M. Nowak and K. Sigmund, *Nature* **364**, 56 (1993).
 - [4] J. Golbeck, *Evolving Strategies for the Prisoners Dilemma*. In *Advances in Intelligent Systems, Fuzzy Systems, and Evolutionary Computation 2002*, p. 299 (2002).
 - [5] J. Laslier, R. Topol, and B. Walliser, *Games and Econ. Behav.* **37**, 340 (2001).
 - [6] R. A. Howard, *Dynamic Programming and Markov Processes* (The MIT Press, Cambridge, Massachusetts, 1960). A. G. Barto *et al.*, in *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, M. Gabriel and J. Moore, Eds. (The MIT Press, Cambridge, Massachusetts, 1991).
 - [7] M. L. Littman, in *Proceedings of the Eleventh International Conference on Machine Learning*, p. 157 (Morgan Kaufmann, San Francisco, CA, 1994).
 - [8] A. W. Beggs, *J. Econ. Th.* **122**, 1 (2005).
 - [9] E. Darmon and R. Waldeck, *Physica A* **355**, 119 (2005).
 - [10] I. Erev and A. E. Roth, *Amer. Econ. Rev.* **88**, 848 (1998).
 - [11] R. Bush and F. Mosteller, *Stochastic Models of Learning* (John Wiley & Son, New York 1955).
 - [12] G. Ódor, *Rev. Mod. Phys.* **76**, 663 (2004). H. Hinrichsen, *Adv. Phys.* **49**, 815 (2000).
 - [13] Ch. Hauert and G. Szabó, *Am. J. Phys.* **73**, 405 (2005).
 - [14] D. Phan, R. Waldeck, M. B. Gordon, and J.-P. Nadal, *Adoption and cooperation in communities: mixed equilibrium in polymorphic populations*. In *proceedings of Wehia05* (University of Essex, United Kingdom, 2005). M. B. Gordon, D. Phan, R. Waldeck, and J. P. Nadal. *Cooperation and free-riding with moral costs*. In *Adv. Cogn. Econ.*, NBU series in Cognitive Science, p. 294 (Sofia NBU Press, 2005).
 - [15] J. Miękisz, in *Lecture notes of CIME Summer School "From a Microscopic to a Macroscopic Description of Complex Systems"* (Będlewo, Poland, 2006).
 - [16] J. M. Baldwin, *Am. Nat.* **30**, 441 (1896).
 - [17] P. Hingston and G. Kendall, *Learning versus Evolution in Iterated Prisoner's Dilemma*, in *Proceedings of Congress on Evolutionary Computation 2004 (CEC'04)*, Portland, Oregon, p. 364 (IEEE, Piscataway NJ, 2004).
 - [18] R. Suzuki and T. Arita, in *Proceedings of 7th International Conference on Neural Information Processing*, p.738 (Taejon, Korea, 2000).
 - [19] G. T. Fechner, *Elemente der Psychophysik* (Breitkopf und Hartel, Leipzig, 1860). S. S. Stevens, *Psychophysics: Introduction to its Perceptual, Neural and Social Prospects* (Wiley, New York, 1975).
 - [20] M. Copelli, A. C. Roque, R. F. Oliveira, and O. Kinouchi, *Phys. Rev. E* **65**, 060901 (2002). O. Kinouchi and M. Copelli, *Nature Physics* **2**, 348 (2006).